DOCUMENT RESUME

ED 366 652 TM 021 076

AUTHOR	Katz, Irvin R.; And Others
TITLE	Extending the Rule Space Model to a Semantically-Rich Domain: Diagnostic Assessment in Architecture.
INSTITUTION	Educational Testing Service, Princeton, N.J.
SPONS AGENCY	Office of Naval Research, Arlington, VA. Cognitive and Neural Sciences Div.
REPORT NO	RR-93-42-ONR
PUB DATE	Oct 93
CONTRACT	N00014-90-J-1307
NOTE	43p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Architects; *Architecture; Classification; *Cognitive Processes; Educational Assessment; *Educational Diagnosis; Item Analysis; Item Response Theory; *Knowledge Level; Problem Solving; Semantics; *Test Construction; Test Items
IDENTIFIERS	Domain Analysis; *Rule Space Model

ABSTRACT

This paper presents a technique for applying the Rule Space Model of cognitive diagnosis (Tatsuoka, 1983) to assessment in a semantically rich domain. Responses of 122 architects to 22 architecture test items developed to assess a range of architectural knowledge were analyzed using Rule Space. Verbal protocol analysis guided the construction of a model of examinee performance, consisting of processes for constructing an initial representation of an item (labeled "understand"), forming goals and performing actions based on those goals ("solve"), and determining whether goals have been attempted and satisfied ("check"). Item attributes derived from these processes formed the basis for diagnosis. Successful diagnostic classifications were obtained for approximately 65 percent, 90 percent, and 40 percent of examinees based, respectively, on attributes associated with the "understand," "solve," and "check" processes of the problem-solving model. The findings support the effectiveness of Rule Space in a complex domain and suggest directions for developing new architecture items by using attributes particularly effective at distinguishing among examinees of different ability levels. Nine tables and three figures present study data. (Contains 23 references.) (Author/SLD)

TM

RR-93-42-ONR

EXTENDING THE RULE SPACE MODEL TO A SEMANTICALLY-RICH DOMAIN: DIAGNOSTIC ASSESSMENT IN ARCHITECTURE

Irvin R. Katz
Michael E. Martinez
Kathleen M. Sheehan
Kikumi K. Tatsuoka

This research was sponsored in part by the
Cognitive Science Program
Cognitive and Neural Sciences Division
Office of Naval Research, under
Contract No. N00014-90-J-1307
R&T 4421559

Kikumi K. Tatsuoka, Principal Investigator

Educational Testing Service

Princeton, NJ

October 1993

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
8/30/93

3. REPORT TYPE AND DATES COVERED
Final, April 1989 - August 1993

4. TITLE AND SUBTITLE
Extending the Rule Space Model to a Semantically-Rich Domain: Diagnostic Assessment in Architecture.

5. FUNDING NUMBERS
C-N00014-90-J-1307
61153 N
RR 04204-01
R & T 4421559

6. AUTHOR(S)
Irvin R. Katz, Michael E. Martinez, Kathleen Sheehan, and Kikumi K. Tatsuoka

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

8. PERFORMING ORGANIZATION REPORT NUMBER
RR-93-42-ONR

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Cognitive Sciences (1142CS)
Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release.
Distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
This paper presents a technique for applying the Rule Space model of cognitive diagnosis (Tatsuoka, 1983) to assessment in a semantically-rich domain. Responses to 22 architecture test items, developed to assess a range of architectural knowledge, were analyzed using Rule Space. Verbal protocol analysis guided the construction of a model of examinee performance, consisting of processess for constructing an initial representation of an item (labeled understand), forming goals and performing actions based on those goals (solve), and determining whether goals have been attempted and satisfied (check). Item attributes, derived from these processes, formed the basis for diagnosis. Successful diagnostic classifications were obtained for approximately 65%, 90%, and 40% of examinees based, respectively, on attributes associated with the understand, solve, and check processes of the problem-solving model. The findings support the effectiveness of Rule Space in a complex domain and suggest directions for developing new architecture items by using attributes particularly effective at distinguishing among examinees of different ability levels.

14. SUBJECT TERMS
diagnostic assessment; problem solving; architecture; rule space; item attributes; computer-based testing.

15. NUMBER OF PAGES
32

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

Extending the Rule Space Model to a Semantically-Rich Domain: Diagnostic Assessment in Architecture

Irvin R. Katz, Michael E. Martinez, Kathleen Sheehan, and Kikumi K. Tatsuoka
Educational Testing Service
Princeton, NJ 08541

Extending the Rule Space Model to a Semantically-Rich Domain: Diagnostic Assessment in Architecture

Abstract

This paper presents a technique for applying the Rule Space model of cognitive diagnosis (Tatsuoka, 1983) to assessment in a semantically-rich domain. Responses to 22 architecture test items, developed to assess a range of architectural knowledge, were analyzed using Rule Space. Verbal protocol analyses guided the construction of a model of examinee performance, consisting of processes for constructing an initial representation of an item (labeled understand), forming goals and performing actions based on those goals (solve), and determining whether goals have been attempted and satisfied (check). Item attributes, derived from these processes, formed the basis for diagnosis. Our technique extends Rule Space's applicability by defining attributes in terms of item characteristics and the causal relations between characteristics and the problem-solving model.

Data were collected from 122 architects of various ability levels (students, architecture interns, and professional architects). Rule Space successfully classified approximately 65%, 90%, and 40% of examinees based, respectively, on attributes associated with the understand, solve, and check processes of the problem-solving model. The findings support the effectiveness of Rule Space in a complex domain and suggest directions for developing new architecture items by using attributes particularly effective at distinguishing among examinees of different ability levels.

Index terms: diagnostic assessment; problems solving; architecture; rule space; item attributes; computer-based testing

Extending the Rule Space Model to a Complex Domain: Diagnostic Assessment in Architecture

As testing programs begin to employ new forms of assessment, a common goal is to construct tests whose demands are closely related to tasks in the target domain (Wiggins, 1989). While recent research has presented several types of assessment tasks (e.g., simulation) that more accurately capture relevant knowledge and skills, there remains the issue of performance reporting: How can we provide examinees with information beyond scores of overall proficiency, information that captures the richness of knowledge and skills in a domain? In the current work, we employ the Rule Space Model (Tatsuoka, 1983) to generate descriptions of examinee ability that are far richer than those normally derived from large-scale assessment. However, Rule Space has been most successfully applied in the past only to relatively narrow topics in well-defined domains (e.g., mixed number subtraction, single-variable isolation in algebra). This paper presents a technique for applying the Rule Space model of cognitive diagnosis (Tatsuoka, 1983) to a semantically-rich domain in need of more authentic, yet tractable, assessments: architecture.

Architecture Assessment

Current architecture assessments consist primarily of short, verbal multiple-choice questions or complex items that mimic the tasks architects normally encounter in the workplace. Because architecture is a complex domain, individuals' scores on relatively simple, verbal multiple-choice tests do not capture the complexity of the knowledge and skills to be assessed. We address these issues by presenting examinees with figural response test items (Martinez, 1991; in press) and by generating diagnostic profiles of examinees based on their performance using the Rule Space model (Tatsuoka, 1983).

The figural response items used in this study differ from standard multiple-choice items in that examinees must construct their answers and the responses consist of the generation or manipulation of figural material (e.g., graphs, pictures). Figural response items are especially suited to domains that are graphical or pictorial in nature; the domain of architecture is a natural candidate for this form of assessment. The approach of using figural response items for architecture assessment has a number of advantages. First, architecture is a graphical domain; designs are drawn, rather than essays being written. Thus, the figural response format provides a natural way for architects to express their ability. Second, constructed response items may be able to tap skills otherwise inaccessible using the multiple-choice format. Martinez & Katz (1992) showed, for example, that different skills are frequently tapped by figural response items compared with their multiple-choice counterparts.

In this study, the figural response items were computer delivered; a sample item is shown in Figure 1. Each item consists of a stem (top of screen), a diagram, and a set of tools for drawing on or manipulating the diagram. The item in Figure 1 requires examinees to move the structures at the bottom of the screen (library, parking lot, and playground) on to the provided site, subject to the explicit constraints stated in the item stem as well as to the implicit constraints that architects associate with libraries, parking lots, and playgrounds (e.g., a playground should not be adjacent to a parking lot; a parking lot must have street access).

Insert Figure 1 about here

Architecture brings certain challenges to the practice of large scale assessment. First, much of architectural practice requires design, a notoriously complex cognitive skill. The duration of design projects in architecture are typically measured in days or months, not minutes as with the

usual examination item. Also, design tasks do not typically have "right" or "wrong" answers. Rather, a continuum of designs satisfy the constraints of the task to a greater or lesser extent. Further, in the real world, constraints on a design task are not immutable; often the architect may relax certain initially specified constraints that he or she believes would allow for a better design (Goel & Pirolli, 1991). We do not seek to assess design skills directly. Although some of the figural response items present simple design tasks, most were meant to assess architectural knowledge through subsidiary tasks. For example, two items present a diagram of a building and ask the candidate to specify locations of seismic joints. While a corresponding task set for an architect might not be this simple, the task could come up as part of a larger design task in the real world.

Architecture may be classified as a "semantically rich domain" (Simon, 1984) in that skilled performance involves extensive specialized knowledge. Architecture knowledge is usually gained over several years of intense study. This knowledge comes from a variety of disciplines, including civil engineering, physics, history, psychology, construction, and art. This forms a second challenge for architectural assessment. Optimally, assessment will produce similarly rich descriptions of proficiency based on test performance. In the current work, we employ the Rule Space Model (Tatsuoka, 1983) to generate descriptions of examinee ability that are far richer than those normally derived from large-scale assessment.

Our approach, like that of many emerging test theories, blends traditional psychometric approaches with developments in cognitive psychology (Gitomer & Yamamoto, 1991). Some new approaches including Rule Space build on item response theory (IRT), in which individuals and items are ordered along a proficiency continuum (Lord & Novick, 1969). One well-known shortcoming of IRT is that identical estimates of overall proficiency may be derived from radically different response patterns. If information about response patterns could be simplified and preserved, these rich descriptions of performance could be truly diagnostic (Mislevy, 1993).

The Rule Space Model

The Rule Space model provides descriptions of examinee performance that extend beyond raw scores or uni-dimensional IRT estimates of overall proficiency. Items are decomposed into attributes, which represent the latent traits that the items assess. Based on an examinee's pattern of correct and incorrect responses, the Rule Space model infers the most likely combination of attributes the examinee has mastered.

The diagnosis of cognitive errors made by examinees is a pattern classification problem. In this study, the patterns are item response vectors, and the vectors are ones and zeroes indicating correct and incorrect responses, respectively. The response vectors are classified as various correct latent knowledge states. The Rule Space model, developed to solve this classification problem, has three steps: (1) determination of classification groups, (2) formulation of a classification space, and (3) classification of examinees' responses.

Determination of Classification Groups

We assume that each postulated cognitive attribute—declarative knowledge, cognitive processes, solution strategies, and so forth—is tapped by at least one item in the pool. The relationship between these cognitive attributes and the items is expressed by an incidence matrix Q, whose order is the number of cognitive attributes k by the number of items n. If item j involves attribute k, then $Q_{kj} = 1$, otherwise $Q_{kj} = 0$. Each item is therefore characterized by the cognitive attributes required for its solution.

For example, suppose there are three items whose two underlying attributes are denoted A_1 and A_2. Further, suppose A_1 is needed to solve items 1 and 3, and A_2 is required in item 2. Then, the incidence matrix **Q** (2x3) is:

	Items
Attribute A1	1 0 1
Attribute A2	0 1 0

With three items, there are eight possible response vectors:

(0,0,0), (1,0,0), (0,1,0), (0,0,1), (1,1,0), (1,0,1), (0,1,1), (1,1,1).

Given two attributes, there are four possible examinee knowledge states:

State 1. Examinee cannot do A_1, but can do A_2
State 2. Examinee cannot do A_2, but can do A_1
State 3. Examinee cannot do A_1 nor A_2
State 4. Examinee can do A_1 and A_2

There are four ideal response vectors conforming to the four states:

State 1. (0,1,0)
State 2. (1,0,1)
State 3. (0,0,0)
State 4. (1,1,1)

Note that each ideal response vector corresponds to a unique vector of mastered attributes. The remaining possible response vectors—(1,0,0), (0,0,1), (1,1,0), (0,1,1)— do not conform precisely to any of the models. The section entitled Classification of Examinees' Responses discusses Rule Space's treatment of such "non-ideal" response vectors.

Tatsuoka (1991) and Varandi & Tatsuoka (1990) developed an algorithm to produce all possible ideal response patterns, corresponding to all possible latent knowledge states from an incidence matrix **Q**. The number of states is determined from the number of attributes, the number of items, and the degree of attribute nesting. In applying Rule Space to other data sets, the number of latent states has often exceeded 1000.

The Classification Space

In order to preserve continuity with current psychometric theories, the classification space was formulated as a two-dimensional Cartesian product space of the IRT proficiency parameter θ, and an index of the unusualness of an item response pattern ζ, where "unusualness" refers to the degree to which easier items are answered incorrectly and difficult items are answered correctly (Tatsuoka & Linn, 1981; Tatsuoka, 1984; 1985; 1990; Tatsuoka & Tatsuoka, 1987). When an examinee's response vector conforms well to the average performances on the test items, the absolute value of ζ will be nearly zero. When ζ-values of a knowledge state are close to zero, that is, close to the θ-axis, we can expect that many examinees will be diagnosed to have that knowledge state. If the ζ-value associated with a knowledge state is large, positively or

negatively, then we expect that state to be unusual in the sense that few examinees will be diagnosed as having that knowledge state.

Classification of Examinees' Responses

Examinees' performances on test items are not always consistent with their unobservable patterns of attribute mastery. Responses that deviate from an ideal response pattern are assumed to contain random errors or slips. Under the assumption that occurrences of slips on items are independent across items, Tatsuoka & Tatsuoka (1987) showed that the distribution of the number of slips follow a binomial distribution if the slippage probabilities are the same across the items, and follow a compound binomial distribution if the slippage probabilities differ across items.

When the non-ideal response patterns associated with a particular ideal pattern, R, are mapped into the Rule Space (by computing their θ and ζ values), they form a unique subset that swarms around the point (θ_R, ζ_R). The swarm of mapped points in the Rule Space follows approximately a multivariate normal distribution with a centroid of (θ_R, ζ_R), and is called the bug distribution or state distribution associated with response pattern R (Tatsuoka, 1990). When all possible ideal item response patterns are mapped on to the Rule Space, one can apply Bayes' decision rules for determining the minimum errors to classify an examinee's point (θ_X, ζ_X) into one of the possible latent states. More detailed discussions of the classification procedure can be found in Tatsuoka (1990), Tatsuoka & Tatsuoka (1987, 1989), and Sheehan, Tatsuoka, & Lewis (1991).

Applying Rule Space to Architecture Assessment

The items used in this research were intended to assess a wide range of architectural knowledge and skills across several subdisciplines of architecture. Different items required different problem-solving operations. For example, some items required examinees to specify the properties of structural elements while others required the proper arrangement of architectural elements on the computer. The range of operations used across items implied that defining attributes in terms of low-level operations would produce an attribute set with little overlap across items. This would defeat the purpose of the Rule Space. We therefore analyzed the architecture items at a coarser grain, using attributes descriptive of higher-level processing as suggested by a general model of problem solving. This approach required a modification to the procedure used in other Rule Space analyses. We first defined a cognitive model that was general enough to account for problem-solving behavior on all items. Attribute definitions were then based on the model. In the next section, we describe the cognitive model and our procedure for defining item attributes.

The Cognitive Model

Our cognitive model was derived in part from a theory of computer interface use (Lewis & Polson, 1990). This model was chosen because of ostensible similarities between problem solving in user interface evaluation and solution of figural response items. Our adaptation of Lewis and Polson's model was based on verbal protocols from one pilot subject who solved all 22 architecture items[1]. The analysis of protocols from a single subject was not used to produce a definitive cognitive model, but a hypothesized model which would guide us in developing reasonable attributes. The reasonableness of this hypothesized model could, in turn, be supported or falsified by our data.

[1]This pilot subject was not part of the test administration discussed in the next section.

The model consists of processes relevant for constructing an initial representation of the item (i.e., understanding the problem stem and provided diagram), forming goals and performing actions based on those goals (i.e., solving the item), and determining whether goals have been satisfied and if they have been satisfied correctly (i.e., checking each problem solving step and the final answer). The model asserts that these processes exist, but makes no claims as to their order. For example, an examinee might come to a new understanding of a problem after attempting to solve it or after checking an initial, incorrect solution. The processes hypothesized by the model are summarized in Table 1.

Insert Table 1 about here

Understand. The first step in solving any item is to understand what is being asked so that the appropriate knowledge can be invoked. Each figural response item consisted of both a verbal stem and a diagram, the latter of which may contain both graphical and verbal information. Thus, understand processes include: (a) reading and interpreting the verbal stem, (b) scanning and interpreting the diagram, and (c) relating the information in the stem and diagram to one's own knowledge. This processing allows the examinee to form initial goals, and either a plan for solving the item or a set of heuristics. An initial goal might be to apply a strategy learned in the classroom or to invoke a general problem-solving method such as means-ends analysis, in which one chooses at each step an action that will reduce the difference between the current state of the problem and the desired goal state. In specifying the understand processes—read stem, scan diagram, and relate to one's own knowledge—no claims are made as to either the ordering of the processes or the conditions under which they occur. Particular items will be less or more difficult in terms of, say, reading and interpreting the stem, and it is just these sorts of differences which form the basis for the item attribute definitions.

Solve. Once an initial representation of the problem has been built, and the initial goals formed, the examinee must perform the actions that lead to solving the problem. Of course, while solving a problem, an examinee may reformulate or refine an initial representation of an item. The processes involved in solving an item are applied to each goal that has not yet been satisfied. Each of these goals may be elaborated by forming subgoals of the currently active goal or the examinee may perform an action that will satisfy the current goal. An action may be physical, such as drawing a line, or cognitive, such as finding a level area on a contour map. These two processes, elaboration of goals and performance of actions, do not determine precisely how a particular item is solved. Certain questions are left open. For example, which subgoals are formed when a particular goal is elaborated? How does the examinee decide on which actions to perform to satisfy a goal? Answering these questions requires a knowledge of the particular strategies used to solve each item. Whatever strategy an examinee uses (whether problem-specific or general), that strategy will determine which goals are attended to and in what order, and what subgoals are formed.

Check. Once an action has been performed, the results of that action may be evaluated to ensure that the action was performed correctly and that it satisfies the original goal. If both conditions are met, the examinee may mark that goal as finished (perhaps by saying something to the effect of "Okay, that's done"), and proceed to the next unsatisfied goal. Thus, two types of evaluations may occur: monitoring whether an action has been carried out as planned and noting whether it satisfies the original goal.

Attribute Creation

Because the figural response items were designed to assess a wide range of architectural knowledge and skill, defining attributes in terms of the actual steps candidates take in solving the items (the approach used in previous applications of Rule Space) was contra-indicated. Instead, we defined attributes in terms of item characteristics or features. Each item has multiple features and could be classified along several dimensions, but for purposes of attribute creation we identified those features with a potential causal connection to examinee performance. The attributes were defined by identifying features of the items that could be expected either to help or hinder problem-solving. For example, we hypothesized that problem solving would be hindered during the process "scan the provided diagram," if the diagram was a specialized graph (e.g., a topographic map) that would not be understood by all examinees. The 38 attributes identified in the task analysis are listed in Table 2. To illustrate the assignment of attributes to items, Table 3 shows the attributes associated with the "library" item of Figure 1 along with an explanation of why that attribute was assigned.

Each attribute is associated with one or more of the three types of processing (understand, solve, and check), and those assignments are shown in Table 4. The assignment of attributes to process was made by two independent judges with an inter-rater agreement of 88%. Disagreements were settled through discussion between the judges. Two independent judges also determined the subset of elementary cognitive attributes needed to solve each question. The inter-rater reliability for this process was again 88%. As before, disagreements were settled through discussion between the judges.

Insert Tables 2, 3, and 4 Here

Method

Materials and Design

Twenty-two figural response questions were constructed to draw upon skills needed throughout the broad content of an architectural licensing examination. These questions were developed for presentation on a computer with responses made through mouse movements and clicks. The questions were divided into two eleven-item subsets, and each subset was administered to a random half of the available subjects[2].

Subjects

Subjects (N=122) were selected from three status groups: practicing architects (N=34), architecture interns (N=35), and architecture students (N=53). The eleven item responses provided by each subject were scored correct/incorrect and modeled with a two-parameter logistic IRT model. Maximum likelihood estimates of proficiency (θ) were subsequently obtained for each subject. These estimates were used to classify subjects into three equal-sized proficiency groups. The cross-tabulation of status groups and proficiency groups is shown in Table 5.

[2]Subjects solved only eleven of the figural response items because they were also administered a set of complementary multiple-choice items. Time constraints did not permit additional testing. Contrasts between item sets are reported in another study (Martinez & Katz, 1992).

Insert Table 5 about here

Procedure

In groups of six, subjects were given a verbal introduction to the item delivery system. Following that, they each attempted the items individually on a computer. Of the 122 subjects, three subjects generated verbal protocols to gather independent support for the cognitive model. To generate the protocols, the subjects were asked to "think aloud" (Ericsson & Simon, 1984), saying anything that they would normally "say" to themselves as they solved the items.

Rule Space Analyses

Rule Space analyses were conducted separately for each of the three groups of problem-solving attributes identified above. This strategy was chosen for two reasons. One very practical reason is that the combination of attributes made the possible number of knowledge states astronomical for the entire set of 38 attributes, thus the total pool of attributes had to be sub-divided. A second reason was to contrast attribute clusters in their ability to classify examinees.

Rule Space was carried out in two steps: First, the BUGLIB computer program (Varandi & Tatsuoka, 1990) was used to determine the set of all possible latent knowledge states associated with the specified stage; second, the RULESPACE computer program (Tatsuoka, Baille & Sheehan, 1990) was used to classify subjects into one of the knowledge states. Three attempts were made to classify each examinee, one for each of the three problem-solving process types (understand, solve, and check).

Results

Verbal Protocol Results

Our cognitive model postulated that certain processes would be used as a subject solved the architecture items. One way to gather evidence for the model is to show that these processes are sufficient for explaining the verbalizations made by subjects (Ericsson & Simon, 1984). Eight categories of subject verbalizations were defined, one category for each process in the cognitive model and a "miscellaneous" category. These categories were defined through examining verbalizations of the pilot subject as she solved eleven of the items. The sufficiency of the categories was established by attempting to categorize the verbalizations on the remaining eleven items. One rater categorized all of the subject's verbalizations, while another rater independently categorized a portion of the verbalizations. The inter-rater agreement on the portion scored by both raters was 82%. The final categories are shown in Table 6. The verbalizations encoded as miscellaneous include single words or short phrases ("Okay," "Let's see"), statements concerning the computer interface ("I have to click twice"), and statements irrelevant to the task.

Insert Table 6 about here

The categorization scheme was applied to the verbal reports of the three protocol subjects. The cognitive model accounted for 71% of the verbalizations made by subjects; the remaining verbalizations fell into the miscellaneous category. This result suggests that the model adequately

captured subjects' problem-solving performance, and thus supports the validity of the cognitive attributes created from this model.

Rule Space Results

The projection of examinee response data into the two-dimensional Rule Space is presented in Figure 2. Examinees' θ values are plotted along the x-axis; ζ values are plotted along the y-axis. The symbols indicate status group membership. The plot shows that practicing architects are located mostly in the medium to high proficiency region and form a cluster that is distinct from the points plotted for interns and students.

Insert Figure 2 about here

Each examinee's performance was diagnosed three times, once for each of the understand, solve, and check attributes. For each diagnosis, the examinee's point in the rule space was compared to the points corresponding to the set of knowledge states associated with each attribute group. The item/attribute incidence matrices developed for each problem-solving process type determined the number of possible states: 803 for understand, 1208 for solve, and 121 for check. Within each process type, each knowledge state corresponded to a unique combination of mastered attributes and is represented by a unique point in the Rule Space.

The classification results for each of the three types of problem-solving processes are presented in Table 7. Within each process type, the number and percentage of classified examinees is broken down by IRT-proficiency level (low, medium, and high) and status group (student, intern, architect). Two patterns are worth noting. The first is that the solve attributes are the most powerful in classifying subjects across proficiency levels and status groups; in fact, all 41 low-proficiency examinees were classified. The next most powerful set of attributes is understand, followed by check. A second pattern is that, almost uniformly, examinees in the lower proficiency or status groups were more often classified than those in the higher groups. For example, twice the percentage of low-proficiency examinees (61%) than high-proficiency examinees (30%) were classified under check.

Insert Table 7 about here

The low classification rate achieved for the check processes is considered in Figure 3. In this plot, the diamonds stand for latent knowledge states and the boxes indicate the examinees' diagnostic location. The plot shows that the 121 knowledge states deduced from the check incidence matrix do not coincide with the examinees' points. Thus, the attributes defined from the check portion of the model do not capture examinee behavior, suggesting that examinee performance is not greatly differentiated by check processes (or that we need to rework that portion of the model).

Insert Figure 3 here

Attribute Mastery Probabilities

An attribute mastery vector was estimated for each classified examinee. These vectors are composed of zeros and ones, depending on whether the attribute in question was included in the subset of mastered attributes defined for the examinee's state. Attribute mastery patterns were averaged within proficiency and status groups, and analyzed using a repeated measures analysis of variance design, as described in Sheehan, Tatsuoka, and Lewis (1991).[3]

P-values for the analysis of variance F-tests are reported in Table 8. The table provides evidence for three clearly significant effects: proficiency group, attribute, and the attribute by proficiency group interaction. These results are reassuring because they indicate that the attributes associated with each problem-solving stage are differentially difficult and that examinees in different proficiency groups tend to have different attribute mastery profiles.

The results obtained for the status group classification are not as clear-cut. Although the main effect of status group is clearly not significant, the interaction of status group with attribute is marginally significant. This indicates that the average probability of mastery values calculated for some attributes differed among students, interns, and practicing architects, but these differences did not hold up after averaging over all attributes. Thus, on the average, examinees in different status groups did not differ in their mastery of the elementary cognitive skills identified in this study.

Insert Table 8 about here

Table 9 presents the mean probability of mastery values estimated for the solve attributes. The different attribute mastery profiles obtained for low, medium and high proficiency examinees are clearly indicated. The table also shows that attributes differ in discrimination. For example, consider the probabilities listed for the "environment" attribute: On average, low proficiency examinees mastered this attribute with a probability of .47; the corresponding probabilities for medium and high proficiency examinees are .60 and .97, respectively. The varying probabilities obtained for low, medium, and high proficiency examinees indicate that this attribute is highly discriminating. By contrast, the three mean values listed for the "learned procedure" attribute are all very similar. Thus, this attribute is not particularly helpful at discriminating among examinees of different ability levels.

Insert Table 9 about here

Discussion and Conclusions

This study exemplifies how an IRT-based model for estimation of overall proficiency can be combined with the diagnostic classification of examinees. The results of the application of Rule Space were satisfying: We were able to classify a large proportion of examinees, especially those of low and medium ability. In principle, these classifications could be reported back to examinees

[3]A standard analysis of variance design would not have been appropriate for these data because the hypothesis of multisample sphericity—that is, independently observed attributes—is violated. The violation results from the fact that, instead of measuring a single attribute on each examinee, our design involves taking 38 attribute measurements. Thus, non-zero correlations are expected among the attribute measurements associated with a particular examinee.

so that remediation in weak areas could proceed. Traditional psychometrics has served well in discriminating among examinees for selection, placement, or classification on the basis of global estimates of proficiency. Rule Space provides estimates of θ, but also yields information that could serve the interests of the examinee in pin-pointing areas of non-mastery. Of course, applications of the technique described in this paper to other complex domains may require a much larger sample size than was used in the current study. Data from a relatively small number of examinees were sufficient for the goal of this paper, which was to demonstrate and explain a methodology for extending Rule Space.

In addition to diagnosis and estimation of θ, Rule Space provides a framework for comparing a model of task performance to examinees' response data. There are few well-defined methodologies for comparing models to data (but see Polk & Newell, 1991), especially those that can accommodate a great variety of individual differences in examinees' knowledge, skill, and strategy. Model testing proceeds as follows: On the basis of a cognitive model, items are analyzed into their component cognitive attributes. The resulting item/attribute matrix (or matrices) leads to strong predictions about examinees' response patterns. If the (θ, ζ) position of an examinee's response pattern is close to that of an ideal response pattern, that examinee is classified into the knowledge state that the response pattern implies. To the extent that examinees' response patterns can be classified, the analysis provides support for the cognitive model. There are of course limitations to the Rule Space method. We have already noted that sets of attributes processed together are limited in size. As they approach 25 or so, the combinations of attribute profiles makes the possible number of ideal states unmanageable. Consequently, the attributes must be clustered and run separately as in this study.

One contribution of this work is that we have outlined a methodology for applying Rule Space to complex domains. Generally, a limitation of Rule Space is that at the level of fine-grained analysis, the operations needed to solve items in a complex domain may not overlap a great deal. Many attributes might in fact be unique to particular items within the item set. If this is the case, the cognitive attributes must be cast at a higher-level of generality such as item characteristics (e.g., type of diagram presented) or general problem-solving approach needed to solve each item (e.g., recalling a fact versus applying a learned procedure). Given more general attributes, what can we say about an examinee's performance? From a psychological viewpoint, the attributes tell us little about the examinee's cognitive competence. But from an educational standpoint, the attributes provide examinees with just the information they need to improve their performance on subsequent tests. The attributes allow us to say that an examinee has difficulties with items having certain properties. While we may have little information about the examinee's skill at a fine-grained level, the diagnostic reports (which attributes are mastered and which aren't) does tell the examinee what types of problems they should seek out and practice solving, and what components of problem solving need special attention.

Attributes should be based on an independently constructed problem-solving model. Analysis of verbal protocols, performed in this work, serves as one means for constructing and verifying a cognitive model. The model supports attribute creation by showing which aspects of the items would help or hinder problem-solving performance. In contrast to developing a list of attributes intuitively, a cognitive model provides a rich description of each attribute because the meaning of each attribute is derived from its place in the model. Methodologically, this rich attribute description promotes a fuller understanding of what each attribute means and facilitates the assigning of attributes to items.

Another contribution of this work is that we were able to examine the power of attributes to discriminate among examinees of various levels. Knowing which attributes are highly

discriminating has value for the construction of items as well as for the design and sequencing of instruction. Differential relevance of attributes across proficiency groups also sheds light on the nature of expert/novice differences in the domain of interest. Rule Space holds a great deal of value for satisfying the requirements of traditional psychometrics and for diagnosis of individual examinees. Through the use of such models, psychometrics has much to offer to learners and teachers beyond estimates of global proficiency.

References

Ericsson, K. A., & Simon, H. A. (1984). Protocol analysis: Verbal reports as data. Cambridge, MA: MIT Press.

Gitomer, D. H. & Yamamoto, K. (1991). Performance modeling that integrates latent trait and class theory. Journal of Educational Measurement, 28, 173-189.

Goel, V., & Pirolli, P. (1991). The structure of design problem spaces (Report No. DPS-3). Berkeley, CA: University of California at Berkeley.

Lewis, C., & Polson, P. G. (1991). Cognitive walkthroughs: A method for theory-based evaluation of user interfaces. Tutorial notes, ACM Computer-Human Interaction Conference.

Lord, F. M., & Novick, M. R. (1969). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Martinez, M. E. (in press). Problem-solving correlates of new assessment forms in architecture. Applied Measurement in Education.

Martinez, M. E. (1991). A comparison of multiple-choice and constructed figural response items. Journal of Educational Measurement, 28, 131-145.

Martinez, M. E, & Katz, I. R. (1992). Cognitive processing requirements of constructed figural response and multiple-choice items in architecture achievement (Research Report No. RR-92-5). Princeton, NJ: Educational Testing Service. Manuscript also submitted for publication.

Mislevy, R. J. (1993). A framework for studying differences between multiple-choice and free-response test items. In R. E. Bennett & W. C. Ward (Eds.). Construction vs. choice in cognitive measurement. (pp. 75-106). Hillsdale, NJ: Erlbaum.

Polk, T. A, & Newell, A. (1991). Analysis of symbolic parameter models (ASPM): A new model-fitting technique for the cognitive sciences. Unpublished manuscript (presented at the 1991 ONR Contractors' Meeting on Cognitive Diagnosis).

Sheehan, K., Tatsuoka, K. K., and Lewis, C. (1991). A diagnostic classification model for document processing skills. Princeton, NJ: Educational Testing Service.

Simon, H. A. (1984). The sciences of the artificial. Cambridge, MA: MIT Press.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. Journal of Educational Measurement, 20(4), 345-354.

Tatsuoka, K. K. (1984). Analysis of errors in fraction addition and subtraction problems (NIE Final Report). Urbana, IL: University of Illinois, CERL.

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. Journal of Educational Statistics, 12(1), 55-73.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error analysis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. G. Shafto (Eds.), Diagnostic monitoring of skill and knowledge acquisition (pp. 453-488). Hillsdale, NJ: Lawrence Erlbaum Associates.

Tatsuoka, K. K. (1991). Boolean algebra applied to determination of the universal set of knowledge states (Tech. Rep. ONR-91-1). Princeton, NJ: Educational Testing Service.

Tatsuoka, K. K., Baille, R., & Sheehan, K. (1991). RULESPACE [computer program]. Princeton, NJ: Educational Testing Service.

Tatsuoka, K. K, & Linn, R. L. (1981). Indices for detecting unusual item response patterns in personnel testing: Links between direct and item-response-theory approaches. Applied Psychological Measurement, 7(1), 81-96.

Tatsuoka, K. K., & Tatsuoka, M. M. (1987). Bug distribution and pattern classification. Psychometrika, 52, 193-206.

Tatsuoka, M. M., & Tatsuoka, K. K. (1989). Rule space. In Kotz & Johnson (Eds.), Encyclopedia of Statistical Sciences. New York: Wiley.

Varandi, F., & Tatsuoka, K. K. (1990). BUGLIB [computer program]. Princeton, NJ: Educational Testing Service.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, 70, 703-713.

Acknowledgments

This research was supported in part by a grant from the Program Research Planning Council of Educational Testing Service and by Contract No. N00014-90-J-1307 from the Office of Naval Research.

For further information and reprints write to Irvin R. Katz, Mail Stop 10-R, Educational Testing Service, Princeton, NJ 08541

Michael E. Martinez is now at the University of California, Irvine.

Table 1
Problem Solving Model

Attribute Group	Processes
Understanding the Item	Read the item stem Scan the diagram Recall relevant information
Solving the Item	Set subgoals Perform actions
Checking Performance	Is the action correct? Is the current goal completed?

Table 2
Attribute Definitions

Attribute Class	Attribute Name	Description	Relations among Attributes in a Class
Characteristics of Presented Figure	Picture	Presented figure is a sketch of an actual object	The three attributes in this class are mutually exclusive (if an item has one attribute in this class, by definition is does not have another attribute from the same class) and exhaustive (all of the items may be classified as having at least one of the attributes in this class)
	Diagram	Presented figure is an abstract diagram of an object	
	Specialized Diagram	Presented figure is a graph or chart -- a visual representation of some information	
Clarity of General Task	Diagram obvious	Based on just the presented figure, its possible for someone to understand what task the item is asking them to perform. Details regarding the task included in the item stem might still be needed for correct performance of the task.	Mutually exclusive, but not exhaustive
	Own obvious	Based on the presented figure along with some prior knowledge, it's possible for someone to understand what task the item is asking them to perform. Details regarding the task included in the item stem might still be needed for correct performance of the task	

Problem-solving requirements of item	Declarative	Requires knowing particular architectural symbols and definitions for correct solution.	Mutually exclusive and exhaustive
	Learned Procedure	Requires the application of fairly standard, algorithmic procedures that usually would have been learned previously.	
	Discovered Strategy	Requires the application of knowledge or procedures in a novel way. These items are more puzzle-like.	
Content area	Site Design	The item tests knowledge or skills associated with one of the recognized subdisciplines of architecture listed to the left.	Mutually exclusive and exhaustive
	Structural Technology (General)		
	Structural Technology (Lateral Forces)		
	Materials and Methods		
	Construction Documents		

Particular Architectural Features	Identify Street	Correct problem solving requires that the candidate can recognize a street on a site plan.	Neither mutually exclusive nor exhaustive
	Environment	Correct problem solving requires that the candidate knows about constraints due to environmental factors (e.g., weather, earthquakes)	
	Contour Lines	Requires the ability to read and interpret contour lines.	
	Forces	Requires the ability to recognize, interpret, and use force vectors.	
General Problem-solving Approach	Read and Translate	Problem solving goes through cycles of getting information from the problem stem, using that information to generate part of the answer, and then repeating.	Mutually exclusive, but not exhaustive
	Indicate Location of New Feature	Problem solving involves placing given elements into new positions or adding information to the provided diagram.	
Response Method	Move/Rotate	Requires arrangement of provided elements.	Exhaustive, but not mutually exclusive
	Label	Requires selecting which of a provided set of labels should be placed at various indicated points on the diagram.	
	Draw Line	Requires drawing of lines onto provided diagram.	
	Draw Arrow	Requires drawing of arrows onto provided diagram.	

Misleading Characteristics	Stem Incorrect	Without detailed knowledge of an item type, the item's stem suggests an incorrect problem-solving method.	Mutually exclusive, but not exhaustive
	Diagram Incorrect	Without detailed knowledge of an item type or diagram type, the item's provided diagram suggests incorrect problem-solving methods.	
Relation between Stem and Problem-solving	Stem Independent	The item stem provides practically no information that could not be gained either through prior knowledge or through the provided figure.	Stem independent and Stem dependent are mutually exclusive and exhaustive.
	Stem Dependent	Problem-solving is necessarily based on information presented in the item stem. This category is the union of "Initial Info" and "Interim Info" categories.	
	Initial Information in Stem	While the stem information is necessary for correct solution, that information is not directly required during the course of problem solving.	Initial info in stem and Interim info in stem are mutually exclusive and exhaustive across Stem dependent items.
	Interim Information in Stem	The information in the stem is needed a number of times during the course of correct problem-solving.	

Completion Criteria	Own Knowledge Stop	Examinees must use their own knowledge to decide whether they are finished responding to an item (i.e., if the answer is complete). Neither the stem nor the diagram directly supply this information.	Mutually exclusive and exhaustive
	Diagram Stop	The provided diagram indicates whether an answer is complete.	
	Diagram and Own Knowledge Stop	The provided diagram along with some specialized knowledge indicates whether an answer is complete.	
	Stem Stop	Information provided in the stem indicates whether a given answer is complete.	
Number of Correct Responses	One Correct	The item has only one correct answer.	Mutually exclusive and exhaustive
	Few Correct	The item has two or three correct answers, which are variants of one another.	
	Many Correct	The item has several correct answers, some of which may be qualitatively different from others and some of which may be variants on another answer.	

Table 3
Attributes Associated with "Library" Item (Figure 1)

Attribute	Explanation
Specialized Diagram	The provided figure is a site plan, which is an abstract diagram of the actual building site. The site plan diagram contains elements that require specialized knowledge to interpret (e.g., contour lines, property lines, symbols for trees).
Diagram Obvious	Based on the provided elements and the operations available (move, etc.), it is clear that the general procedure for this task is to place the elements somewhere onto the site.
Discovered Strategy	There is no clear, algorithmic procedure for placing the buildings onto the site. The examinees must bring to bear knowledge learned in different situations to the solving of this task.
Site Design	This item presents a prototypical site design task.
Identify Street	Recognizing the street on the site plan is important for correct placement of the parking lot.
Contour Lines	Correctly interpreting the site plan's contour lines is necessary for correct placement of the buildings on the site (e.g., the buildings should not be placed on the steep slope, but on relatively level ground).
Stem Independent	Beyond the general task and the standard "preserve all trees," the stem does not provide any information that is vital to the correct solution of the item.
Many Correct	There are a number of correct solutions to this item, reflecting different arrangements of the buildings on the site.
Move/Rotate	The primary interface operation in this task is moving elements and rotating them to fit better onto the site.
Own Stop	Based on their own knowledge, it is up to the examinees to determine when they are finished responding to the item. Nothing in the stem nor in the diagram provides feedback either on the correctness or completeness of a response.

Table 4
Attribute Assignments to Processing Types

Attribute Class	Attribute	Problem-Solving Process Type		
		Understand	Solve	Check
Characteristics of presented figure	Picture	X		
	Diagram	X		
	Specialized Diagram	X		
Clarity of general task	Diagram Obvious	X	X	
	Own Obvious	X	X	
Problem-solving requirements of item	Learned Algorithm	X	X	
	Declarative	X	X	
	Discovered Strategy	X	X	
Content area	Site Design	X		
	Structural Technology	X		
	Structural Tech. (Lateral Forces)	X		
	Materials and Methods	X		
	Construction Documents	X		
Particular architectural features	Identify Street	X	X	
	Environment	X	X	
	Contour Lines	X	X	
	Forces	X	X	
Relation between stem and problem-solving	Stem Independent	X		
	Stem Dependent	X		
	Initial Info in Stem	X		
	Interim Info. in Stem		X	
Number of correct responses	One Correct			X
	Few Correct		X	X
	Many Correct		X	X
General problem-solving approach	Read and Translate		X	
	Indicate Location of New Feature		X	
Response method	Move/Rotate		X	
	Label		X	
	Draw Line		X	
	Draw Arrow		X	
Completion Criteria	Own Stop			X
	Diagram Stop			X
	Stem Stop			X
	Diagram + Own Stop			X
Misleading Characteristics	Stem Incorrect			X
	Diagram Incorrect			X

Table 5
Distribution of Status Groups by Proficiency

Proficiency	N	Student N	Student Column %	Intern N	Intern Column %	Architect N	Architect Column %
Low	41	27	51	10	29	4	12
Medium	41	17	32	12	34	12	35
High	40	9	17	13	37	18	53
Total	122	53	100	35	100	34	100

Table 6
Protocol Encoding Categories

Understand

Read stem: statements involving the reading of the problem stem. Read statements include verbatim readings of the stem as well as partial reading of the problem stem.

Scan diagram: statements involving the provided diagram. Diagram statements include verbatim readings of verbal information as well as verbal descriptions of information in the diagram (e.g., "lateral forces coming that way").

Relate: statements regarding how the problem or parts of the problem relate to the examinee's own knowledge. Relate statements consist of several types of verbalizations including verbalizations regarding:

- an expectation or the violation of an expectation (e.g., "Normally there would be more lines on this window drawing")
- recognition of the problem (or part of the problem) as of a particular type (e.g., "This is a site vignette," "this is a perspective drawing")
- predictions as to the difficulty of the problem (e.g., "this will take a while")
- the definitions or ambiguity of sections of the problem (e.g., "is it an awning or a hopper?", "most sheathing I know of is...")

Solve

Goal: stating an intent or future action. Goal statements are often stated in the future tense or in terms of "should be."

Perform: statements regarding the performance of an action. Perform statements are usually stated in the present or "continuing" tense (e.g., "that dips here"). Perform statements relate only to physical actions such as moving a block on the screen or locating a particular item in the diagram (for the latter, e.g., "this is a flat area"). It may be difficult to distinguish between goal and perform statements.

Check

Evaluate-correct: statements regarding the correctness of a performed action or the result of that action (e.g., the location of a placed object). Evaluate-correct statements should only refer to the examinee's own actions or answers, not to the problem itself. These statements may either reflect judging the correctness of an action (e.g., "is that right?") or reflect the outcomes of a judgment (e.g., "that isn't what I wanted to do").

Evaluate-complete: statements suggesting that some action or goal has been completed. As with evaluate-correct statements, evaluate-complete statements include verbalizations judging if something has been finished (e.g., "is there anything else to be done?") as well as verbalizations concerning the results of such judgments (e.g., "that's it," "that was easy").

Table 7
Classification Results for Subjects Grouped by Proficiency and Status

Group	N	Problem-Solving Process Type					
		Understand		Solve		Check	
		No. Classified	%	No. Classified	%	No. Classified	%
Proficiency							
Low	41	32	78	41	100	25	61
Medium	41	29	71	40	98	13	32
High	40	20	50	33	83	12	30
Status							
Student	53	37	70	52	98	25	47
Intern	35	23	66	32	91	11	31
Architect	34	21	62	30	88	14	41
Total	122	81	66	114	93	50	41

Table 8
Analysis of Variance Results for Attribute Mastery Data

Group	Problem-Solving Process Type		
	Understand p-value	Solve p-value	Check p-value
Between Subjects			
Proficiency	.0001	.0001	.0001
Status	.5621	.1948	.3433
Proficiency x Status	.1343	.4707	.7231
Within Subjects[a]			
Attribute	.0001	.0001	.0001
Attribute x Proficiency	.0013	.0001	.0130
Attribute x Status	.0885	.0874	.4287
Attr. x Prof. x Status	.4743	.0535	.1029

[a] p-values for within-subject effects were calculated using Wilks' Lambda.

Table 9
Attribute Mastery Probabilities for Solve

Attribute	Proficiency			Overall Mean
	Low	Medium	High	
Many correct	.35	.28	.35	.34
Draw Arrow	.44	.38	.33	.38
Move/Rotate	.31	.39	.61	.44
Label	.43	.66	.88	.66
Environment	.47	.60	.97	.68
Contour Lines	.61	.78	.83	.74
Forces	.70	.64	.88	.74
Identify Street	.75	.76	.84	.78
Interim Info	.67	.82	.92	.80
Diagram Obvious	.70	.76	.94	.80
Own Obvious	.81	.83	.86	.83
Few Correct	.66	.91	.98	.85
Discovered Strategy	.79	.89	.98	.89
Ind. Location	.75	1.00	1.00	.92
Read + Translate	.86	.98	.98	.94
Declarative	.87	.97	1.00	.95
Learned Algorithm	.92	.97	.98	.96
Stem Independent	.89	.99	1.00	.96
Stem Dependent	.93	1.00	1.00	.98
Draw Line	.98	1.00	1.00	.99
Overall Mean	.69	.78	.87	.78

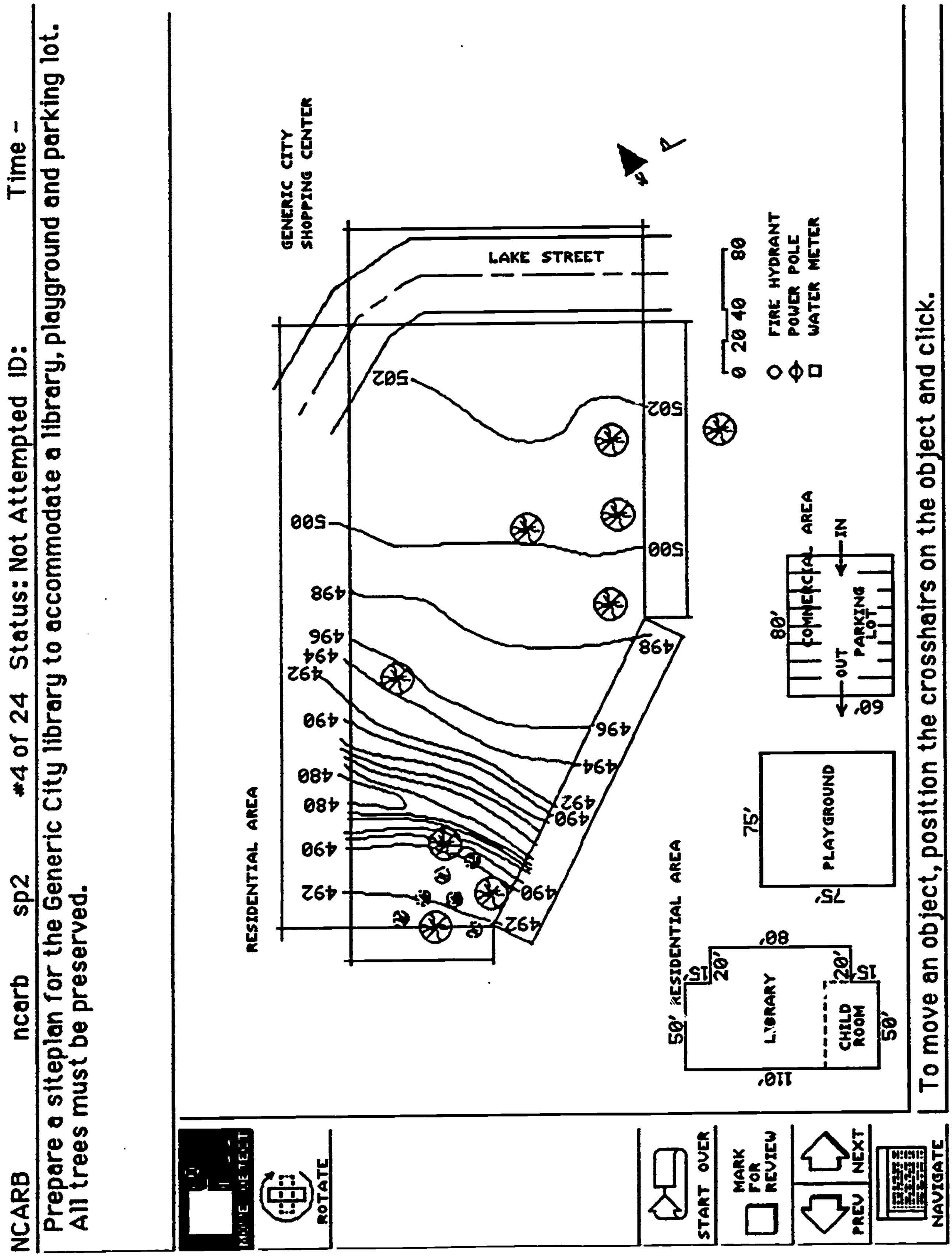

Figure 1. Sample figural response item.

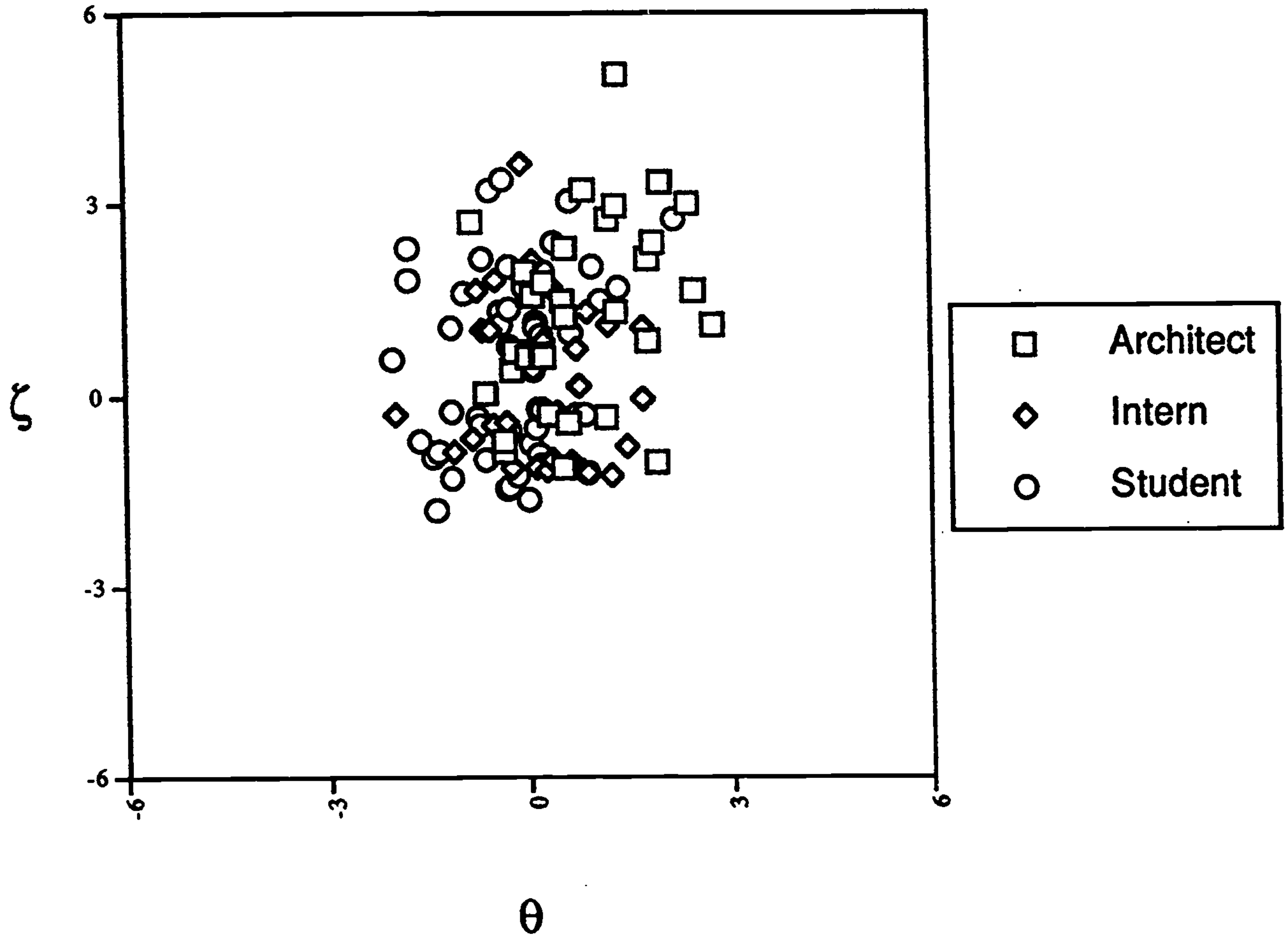

Figure 2. Projection of examinee response data onto (θ, ζ) space.

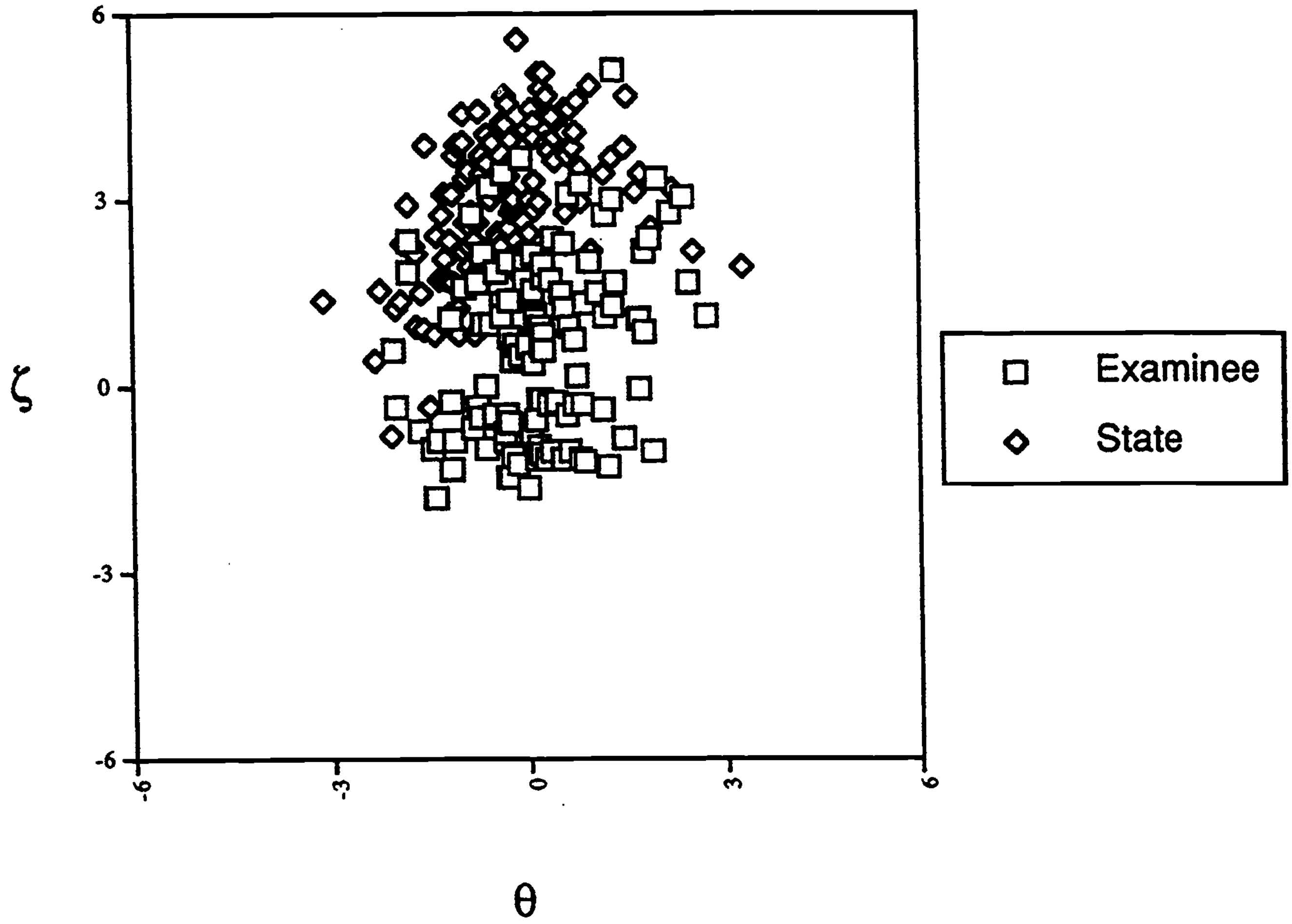

Figure 3. Knowledge states and examinee data for check matrix.

Dr Terry Ackerman
Educational Psychology
260C Education Bldg
University of Illinois
Champaign IL 61801

Dr Terry Allard
Code 3422
Office of Naval Research
800 N Quincy St
Arlington VA 22217-5660

Dr Nancy Allen
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Gregory Anrig
Educational Testing Service
Mail Stop 14-C
Princeton NJ 08541

Dr Phipps Arabie
Graduate School of Management
Rutgers University
92 New Street
Newark NJ 07102-1895

Dr Isaac I Bejar
Educational Testing Service
Mail Stop 11-R
Princeton NJ 08541

Dr William O Berry
Director
Life and Environmental Sciences
AFOSR/NL N1
Bldg 410
Bolling AFB DC 20332-6448

Dr Thomas G Bever
Department of Psychology
University of Rochester
River Station
Rochester NY 14627

Dr Menucha Birenbaum
Educational Testing Service
Princeton NJ 08541

Dr Bruce Bloxom
Defense Manpower Data Center
99 Pacific St
Suite 155A
Monterey CA 93943-3231

Dr Gwyneth Boodoo
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Richard L Branch
HQ USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago IL 60064

Dr Robert Brennan
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr David V Budescu
Department of Psychology
University of Haifa
Mount Carmel Haifa 31999
ISRAEL

Dr Gregory Candell
CTB/MacMillan/McGraw-Hill
2500 Garden Road
Monterey CA 93940

Dr Paul R Chatelier
PERCEPTRONICS
1911 North Ft Myer Drive
Suite 1100
Arlington VA 22209

Dr Susan Chipman
Cognitive Science Program
Office of Naval Research
800 North Quincy Street
Code 3422
Arlington VA 22217-5660

Dr Raymond E Christal
UES LAMP Science Advisor
AL/HRMIL
Brooks AFB TX 78235

Dr Norman Cliff
Department of Psychology
University of Southern California
Los Angeles CA 90089-1061

Director
Life Sciences
Code 3420
Office of Naval Research
Arlington VA 22217-5660

Commanding Officer
Naval Research Laboratory
Code 4827
Washington DC 20375-5000

Dr John M Cornwell
Department of Psychology
I/O Psychology Program
Tulane University
New Orleans LA 70118

Dr William Crano
Department of Psychology
Texas A&M University
College Station TX 77843

Dr Linda Curran
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn VA 22209

Professor Clément Dassa
Faculté des sciences de l'éducation
Département d'études en éducation
et d'administration de l'éducation
CP 6128 succursale A
Montéal Québec
CANADA H3C 3J7

Dr Timothy Davey
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Charles E Davis
Educational Testing Service
Mail Stop 16-T
Princeton NJ 08541

Dr Ralph J DeAyala
Meas Stat and Eval
Benjamin Bldg Room 1230F
University of Maryland
College Park MD 20742

Dr Sharon Derry
Florida State University
Department of Psychology
Tallahassee FL 32306

Hei-Ki Dong
BELLCORE
6 Corporate Place
RM: PYA-1K207
PO Box 1320
Piscataway NJ 08855-1320

Dr Neil Dorans
Educational Testing Service
Mail Stop 07-E
Princeton NJ 08541

Dr Fritz Drasgow
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Defense Tech Information Center
Cameron Station Bldg 5
Alexandria VA 22314
(2 Copies)

Dr Richard Duran
Graduate School of Education
University of California
Santa Barbara CA 93106

Dr Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence KS 66045

Dr George Engelhard Jr
Division of Educational Studies
Emory University
210 Fishburne Bldg
Atlanta GA 30322

ERIC Facility-Acquisitions
2440 Research Blvd
Suite 550
Rockville MD 20850-3238

Dr Marshall J Farr
Farr-Sight Co
2520 North Vernon Street
Arlington VA 22207

Dr Leonard Feldt
Lindquist Center for Measurement
University of Iowa
Iowa City IA 52242

Dr Richard L Ferguson
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Gerhard Fischer
Liebiggasse 5
A 1010 Vienna
AUSTRIA

Dr Myron Fischl
US Army Headquarters
DAPE-HR
The Pentagon
Washington DC 20310-0300

Mr Paul Foley
Navy Personnel R&D Center
San Diego CA 92152-6800

Chair
Department of Computer Science
George Mason University
Fairfax VA 22030

Dr Robert D Gibbons
University of Illinois at Chicago
NPI 909A M/C 913
912 South Wood Street
Chicago IL 60612

Dr Janice Gifford
University of Massachusetts
School of Education
Amherst MA 01003

Dr Robert Glaser
Learning Res & Development Cntr
University of Pittsburgh
3939 O'Hara Street
Pittsburgh PA 15260

Dr Susan R Goldman
Peabody College
Box 45
Vanderbilt University
Nashville TN 37203

Dr Timothy Goldsmith
Department of Psychology
University of New Mexico
Albuquerque NM 87131

Dr Sherrie Gott
AFHRL/MOMJ
Brooks AFB TX 78235-5601

Dr Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore MD 21218

Professor Edward Haertel
School of Education
Stanford University
Stanford CA 94305-3096

Dr Ronald K Hambleton
University of Massachusetts
Lab of Psychom & Eval Res
Hills South Room 152
Amherst MA 01003

Dr Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign IL 61820

Dr Patrick R Harrison
Computer Science Department
US Naval Academy
Annapolis MD 21402-5002

Ms Rebecca Hetter
Navy Personnel R&D Center
Code 13
San Diego CA 92152-6800

Dr Thomas M Hirsch
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Professor Paul W Holland
Div of Educ Psych & Quant
Methods Prog
Graduate School of Education
4511 Tolman Hall
University of California-Berkeley
Berkeley CA 94720

Professor Lutz F Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Ms Julia S Hough
Cambridge University Press
40 West 20th Street
New York NY 10011

Dr William Howell
Chief Scientist
AFHRL/CA
Brooks AFB TX 78235-5601

Dr Huynh Huynh
College of Education
University of South Carolina
Columbia SC 29208

Dr Martin J Ippel
Center for the Study of
Education and Instruction
Leiden University
PO Box 9555
2300 RB Leiden
THE NETHERLANDS

Dr Robert Jannarone
Elec and Computer Eng Dept
University of South Carolina
Columbia SC 29208

Dr Kumar Joag-dev
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Professor Douglas H Jones
Grad Sch of Management
Rutgers The State University NJ
Newark NJ 07102

Dr Brian Junker
Carnegie-Mellon University
Department of Statistics
Pittsburgh PA 15213

Dr Marcel Just
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh PA 15213

Dr J L Kaiwi
Code 442/JK
Naval Ocean Systems Center
San Diego CA 92152-5000

Dr Michael Kaplan
Office of Basic Research
US Army Research Institute
5001 Eisenhower Avenue
Alexandria VA 22333-5600

Dr Jeremy Kilpatrick
Dept of Mathematics Education
105 Aderhold Hall
University of Georgia
Athens GA 30602

Ms Hae-Rim Kim
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr. Jwa-keun Kim
Department of Psychology
Middle Tennessee State University
Murfreesboro TN 37132

Dr Sung-Hoon Kim
KEDI
92-6 Umyeon-Dong
Seocho-Gu
Seoul
SOUTH KOREA

Dr G Gage Kingsbury
Portland Public Schools
Res & Eval Department
501 North Dixon Street
PO Box 3107
Portland OR 97209-3107

Dr William Koch
Box 7246
Meas & Eval Center
University of Texas-Austin
Austin TX 78703

Dr James Kraatz
Computer-based Education
Research Laboratory
University of Illinois
Urbana IL 61801

Dr Patrick Kyllonen
AFHRL/MOEL
Brooks AFB TX 78235

Ms Carolyn Laney
1515 Spencerville Rod
Spencerville MD 20868

Richard Lanterman
Commandant (G-PWP)
US Coast Guard
2100 Second Street SW
Washington DC 20593-0001

Dr Michael Levine
Educational Psychology
210 Education Building
1310 South Sixth Street
Univ of IL at Urbana-Champaign
Champaign IL 61820-6990

Dr Charles Lewis
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541-0001

Mr Hsin-hung Li
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Library
Naval Training Systems Center
12350 Research Parkway
Orlando FL 32826-3224

Dr Marcia C Linn
Graduate School of Education
EMST
Tolman Hall
University of California
Berkeley CA 94720

Dr Robert L Linn
Campus Box 249
University of Colorado
Boulder CO 80309-0249

Logicon Inc (Attn: Library)
Tactical & Training Systems Div
PO Box 85158
San Diego CA 92138-5158

Dr Richard Luecht
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr George B. Macready
Dept of Meas Stat & Eval
College of Education
University of Maryland
College Park MD 20742

Dr Evans Mandes
George Mason University
4400 University Drive
Fairfax VA 22030

Dr Paul Mayberry
Center for Naval Analysis
4401 Ford Avenue
PO Box 16268
Alexandria VA 22302-0268

Dr James R McBride
HumRRO
6430 Elmhurst Drive
San Diego CA 92120

Mr Christopher McCusker
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Dr Robert McKinley
Educational Testing Service
Princeton NJ 08541

Dr Joseph McLachlan
Navy Pers Res & Dev Cntr
Code 14
San Diego CA 92152-6800

Alan Mead
c/o Dr Michael Levine
Educational Psychology
210 Education Bldg
University of Illinois
Champaign IL 61801

Dr Timothy Miller
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Robert Mislevy
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Ivo Molenar
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Grote Kruisstraat 2/1
9712 TS Groningen
The NETHERLANDS

Dr Eiji Muraki
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Ratna Nandakumar
Educational Studies
Willard Hall Room 213E
University of Delaware
Newark DE 19716

Acad Prog & Research Branch
Naval Tech Training Command
Code N-62
NAS Memphis (75)
Millington TN 30854

Dr W Alan Nicewander
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Head
Personnel Systems Department
NPRDC (Code 12)
San Diego CA 92152-6800

Director
Training Systems Department
NPRDC (Code 14)
San Diego CA 92152-6800

Library NPRDC
Code 041
San Diego CA 92152-6800

Librarian
Naval Cntr for Applied Research
in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington DC 20375-5000

Office of Naval Research
Code 3422
800 N Quincy Street
Arlington VA 22217-5660
(6 Copies)

ONR Resident Representative
New York City
33 Third Avenue - Lower Level
New York NY 10003-9998

Special Asst for Res Management
Chief of Naval Personnel
(PERS-O1JT)
Department of the Navy
Washington DC 20350-2000

Dr Judith Orasanu
NASA Ames Research Center
Mail Stop 239-1
Moffett Field CA 94035

Dr Peter J Pashley
Law School Admission Services
PO Box 40
Newtown PA 18940-0040

Wayne M Patience
American Council on Education
GED Testing Service Suite 20
One Dupont Circle NW
Washington DC 20036

Dept of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey CA 93943-5026

Dr Peter Pirolli
School of Education
University of California
Berkeley CA 94720

Dr Mark D Reckase
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Mr Steve Reise
Department of Psychology
University of California
Riverside CA 92521

Mr Louis Roussos
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr Donald Rubin
Statistics Department
Science Center Room 608
1 Oxford Street
Harvard University
Cambridge MA 02138

Dr Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg
Knoxville TN 37966-0900

Dr Mary Schratz
4100 Parkside
Carlsbad CA 92008

Mr Robert Semmes
N218 Elliott Hall
Department of Psychology
University of Minnesota
Minneapolis MN 55455-0344

Dr Valerie L Shalin
Dept of Industrial Engineering
State University of New York
342 Lawrence D Bell Hall
Buffalo NY 14260

Mr Richard J Shavelson
Graduate School of Education
University of California
Santa Barbara CA 93106

Kathleen Sheehan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr Randall Shumaker
Naval Research Laboratory
Code 5500
4555 Overlook Avenue SW
Washington DC 20375-5000

Dr Judy Spray
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Martha Stocking
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St
Champaign IL 61820

Dr Kikumi Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr David Thissen
Psychometric Laboratory
CB# 3270 Davie Hall
University of North Carolina
Chapel Hill NC 27599-3270

Mr Thomas J Thomas
Federal Express Corporation
Human Resource Development
3035 Director Row Suite 501
Memphis TN 38131

Mr Gary Thomasson
University of Illinois
Educational Psychology
Champaign IL 61820

Dr Howard Wainer
Educational Testing Service
15-T Rosedale Road
Princeton NJ 08541

Elizabeth Wald
Office of Naval Technology
Code 227
800 North Quincy Street
Arlington VA 22217-5000

Dr Michael T Waller
Univ of Wisconsin-Milwaukee
Educ Psychology Department
Box 413
Milwaukee WI 53201

Dr Ming-Mei Wang
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Thomas A Warm
FAA Academy
PO Box 25082
Oklahoma City OK 73125

Dr David J Weiss
N660 Elliott Hall
University of Minnesota
75 E River Road
Minneapolis MN 55455-0344

Dr Douglas Wetzel
Code 15
Navy Personnel R&D Center
San Diego CA 92152-6800

German Military Representative
Personalstammamt
Koelner Str 262
D-5000 Koeln 90
WEST GERMANY

Dr David Wiley
Sch of Educ and Social Policy
Northwestern University
Evanston IL 60208

Dr Bruce Williams
Dept of Educational Psychology
University of Illinois
Urbana IL 61801

Dr Mark Wilson
School of Education
University of California
Berkeley CA 94720

Dr Eugene Winograd
Department of Psychology
Emory University
Atlanta GA 30322

Dr Martin F Wiskoff
PERSEREC
99 Pacific Street
Suite 4556
Monterey CA 93940

Mr John H Wolfe
Navy Personnel R&D Center
San Diego CA 92152-6800

Dr Kentaro Yamamoto
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Duanli Yan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey CA 93940

Dr Joseph L Young
National Science Foundation
Room 320
1800 G Street NW
Washington DC 20550